Mind the Gap

Blinder-Oaxaca and Non-Linear Decompositions

So, you've run a regression. Congratulations! You've successfully established that Group A earns significantly more money, has better health outcomes, or is more likely to get approved for a mortgage than Group B. The coefficient on your Group_A dummy variable has three stars next to it, and your p-value has more zeros than a Silicon Valley bank account.

But your thesis advisor, sitting across from you with a cup of artisanal matcha, asks the dreaded question: *"Okay, but **why** is there a gap?"* Is the wage gap between Group A and Group B because Group A has more education? Or is it because the labor market *rewards* Group A's education more than Group B's? Is the gap driven by actual differences in résumés (what economists call "endowments"), or is it driven by systemic bias, discrimination, and unobservable vibes (what statisticians awkwardly call "differences in coefficients")?

Welcome to the world of **Decomposition Methods**. In this guide we dive into the legendary **Blinder-Oaxaca (BO) decomposition** (developed independently by Alan Blinder and Ronald Oaxaca in 1973). Then, because it's the 21st century and we often deal with binary outcomes (like "Did they click 'Subscribe' or not?"), we'll venture into the Matrix to learn about the **Fairlie non-linear decomposition** for logit and probit models.

Grab your iced coffee. Let's deconstruct some gaps.

---

1. The OG: The Linear Blinder-Oaxaca Decomposition

Start with the classic setup. You have a continuous outcome variable Y (say, the natural log of hourly wages). You have two groups: Group A (Jedi) and Group B (Sith). You observe that, on average, Jedi get paid more than Sith. The raw mean gap is:

$$\Delta \bar{Y} = \bar{Y}_A - \bar{Y}_B$$

In a linear regression the mean of the dependent variable equals exactly the mean of the predicted values. The mean outcome for each group can therefore be written as the product of average characteristics ($\bar{X}$ vector) and estimated coefficients (β vector):

$$\bar{Y}_A = \bar{X}_A \hat{\beta}_A \qquad \bar{Y}_B = \bar{X}_B \hat{\beta}_B$$

So the gap can be rewritten as:

$$\Delta \bar{Y} = \bar{X}_A \hat{\beta}_A - \bar{X}_B \hat{\beta}_B$$

This equation is absolutely true, but it's about as helpful as your ex texting you "we need to talk." It doesn't tell us *what* is driving the gap. To figure that out, we need to do some algebraic gymnastics.

The Three-Fold Decomposition

What if we added and subtracted the exact same term, $\bar{X}_B\beta_A$, into our equation? (This is the mathematical equivalent of doing nothing, but making it look fancy.)

$$\Delta \bar{Y} = \bar{X}_A \hat{\beta}_A - \bar{X}_B \hat{\beta}_A + \bar{X}_B \hat{\beta}_A - \bar{X}_B \hat{\beta}_B$$

Adding and subtracting $\bar{X}_A\beta_B$ and juggling these terms around gives the **Three-Fold Decomposition**:

$$\Delta \bar{Y} = \underbrace{(\bar{X}_A - \bar{X}_B)' \hat{\beta}_B}_{E\,:\,\text{Endowments}} + \underbrace{\bar{X}_B' (\hat{\beta}_A - \hat{\beta}_B)}_{C\,:\,\text{Coefficients}} + \underbrace{(\bar{X}_A - \bar{X}_B)' (\hat{\beta}_A - \hat{\beta}_B)}_{I\,:\,\text{Interaction}}$$

E	C	I
ENDOWMENTS	**COEFFICIENTS**	**INTERACTION**
Different résumé characteristics evaluated at Group B pay rates	Different rewards for the same characteristics — potential discrimination	Groups differ in both characteristics and pay rates simultaneously

• **Endowments (E):** The part of the gap explained by the fact that Jedi and Sith have different résumés. Maybe Jedi have more years of lightsaber training. Evaluated using the Sith's baseline pay scale (β_B). Put simply: *How much more would Sith earn if they had Jedi-level training, assuming they are still paid like Sith?*

• **Coefficients (C):** The part of the gap explained by the fact that the market *rewards* traits differently based on who you are. This is often where "discrimination" lives. Put simply: *How much more would Sith earn if their current résumés were suddenly rewarded at Jedi pay rates?*

• **Interaction (I):** The leftover math soup. It accounts for the fact that Jedi and Sith have different résumés *and* are paid differently for them at the exact same time.

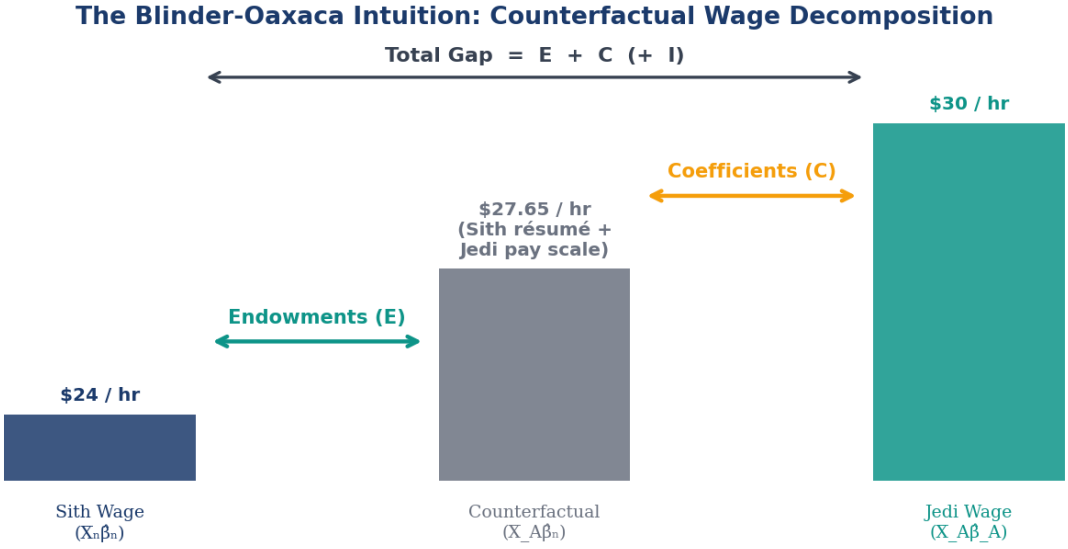

Figure 1 — The BO decomposition as a counterfactual wage bridge. The "counterfactual" bar shows what Sith would earn with Jedi characteristics but still paid according to the Sith pay scale. The gap from Sith to the counterfactual is the Endowments component (E); the gap from the counterfactual to Jedi is the Coefficients component (C).

The Two-Fold Decomposition and the "Index Number" Problem

The three-fold decomposition is mathematically elegant, but that interaction term is super annoying to explain at academic conferences. Because of this, applied researchers usually prefer the **Two-Fold Decomposition**.

The two-fold approach assumes there is some hypothetical, non-discriminatory, utopian reference set of coefficients — call it β*. In a world without bias, everyone would be paid according to β*. The gap then splits cleanly into two pieces:

$$\Delta \bar{Y} = \underbrace{(\bar{X}_A - \bar{X}_B)'\hat{\beta}^*}_{\text{Explained}} + \underbrace{\left[\bar{X}'_A(\hat{\beta}_A - \hat{\beta}^*) + \bar{X}'_B(\hat{\beta}^* - \hat{\beta}_B)\right]}_{\text{Unexplained}}$$

But wait… what the heck is β*? This is the *Index Number Problem*. Who decides what the "fair" market rate is?

- **Oaxaca (1973)** suggested using Group A's coefficients ($\beta^* = \beta_A$) if we assume there's no positive discrimination *for* Group A, only negative discrimination *against* Group B.
- **Reimers (1983)** said, "Let's be diplomatic and split it 50/50!" ($\beta^* = 0.5\beta_A + 0.5\beta_B$).
- **Cotton (1988)** pointed out that groups are rarely the same size, so we should weight the coefficients by sample size.
- **Neumark (1988) and Oaxaca & Ransom (1994)** hit us with the ultimate solution: just pool the data. Run a single regression on everyone combined to get β*.

> ⚠ **Critical: The Pooled-Model Dummy Variable Rule**
>
> If you use Neumark's pooled model approach, **you MUST include a dummy variable for group membership in the pooled regression.** If you don't, any unexplained discrimination will bleed into the coefficients of your explanatory variables (like education) because of omitted variable bias. If Jedi are better educated *and* favored by bias, a pooled model without a Jedi dummy will artificially inflate the return to education, overstating the "Explained" portion of your gap.

The Categorical Variable Trap

Imagine you include a categorical variable in your model — say, “Undergraduate Major,” with “History” as your omitted baseline reference category.

Here is a horrifying truth: in a Blinder-Oaxaca decomposition, the “Unexplained” (Coefficients) component for your categorical variables will **change depending on which category you omit.** If you make “Computer Science” the baseline instead of “History,” the amount attributed to major-specific coefficient differences shifts, trading off with the model’s intercept.

To fix this, researchers like Gardeazabal & Ugidos (2004) and Yun (2005) popularized the **Deviation Contrast Transform**: instead of expressing coefficients as deviations from an arbitrary baseline, express them as deviations from the *grand mean* of all categories. This forces category coefficients to sum to zero, making results completely invariant to the choice of omitted category. Most modern software (the *oaxaca* package in R or Stata) does this automatically — but it’s crucial to know *why* it’s happening so you don’t panic when looking at the raw output.

2. Structural vs. Reduced Form: Peeking Inside the Black Box

In his original 1973 paper, Alan Blinder made a brilliant point that is often forgotten by modern researchers. He ran *two* different types of wage regressions: Structural and Reduced Form.

Structural Equation

Includes *everything happening right now* — education, occupation, union membership, tenure, etc. When Blinder ran this for White vs. Black males, roughly **60 %** of the wage gap was explained by Endowments (Black men had less education, fewer union jobs, etc.) and **40 %** was Unexplained (direct wage discrimination).

Reduced Form Equation

Blinder argued that "Occupation" and "Education" aren't actually exogenous. Systemic racism affects what schools you can access and what occupations you can enter *long before you ask for a paycheck*. So he ran a second regression predicting wages based only on the **circumstances of birth** (parents' education, growing up in the South, rural vs. urban childhood).

The Plot Twist: When using the Reduced Form, Endowments explained only **30 %** of the gap — the Unexplained portion skyrocketed to **70 %**.

Key Takeaway

When you use a structural model, you are only measuring discrimination *at the point of pay*. You are treating occupational segregation and unequal access to education as "objective differences in résumés." If you want to capture the full, systemic picture, you need to think carefully about which variables are truly exogenous.

3. Entering the Matrix: The Fairlie Non-Linear Decomposition

Fast forward to 1999. The world is worried about Y2K, *The Matrix* is in theaters, and Robert Fairlie is looking at binary outcomes.

Up until now, the Blinder-Oaxaca decomposition was strictly for OLS (linear models). But what if your outcome is binary? What if Y is "Started a Business" (1 = Yes, 0 = No)?

You *could* just run a Linear Probability Model (LPM) and slap a standard Blinder-Oaxaca on it. But LPMs have a fatal flaw: they assume the effect of moving from $10,000 in wealth to $20,000 is the exact same as moving from $1,000,000 to $1,010,000. In reality, probabilities are bounded between 0 and 1. They follow an

S-curve (like a Logit or Probit). If you use a linear BO decomposition when there are massive group differences in an independent variable, the math blows up.

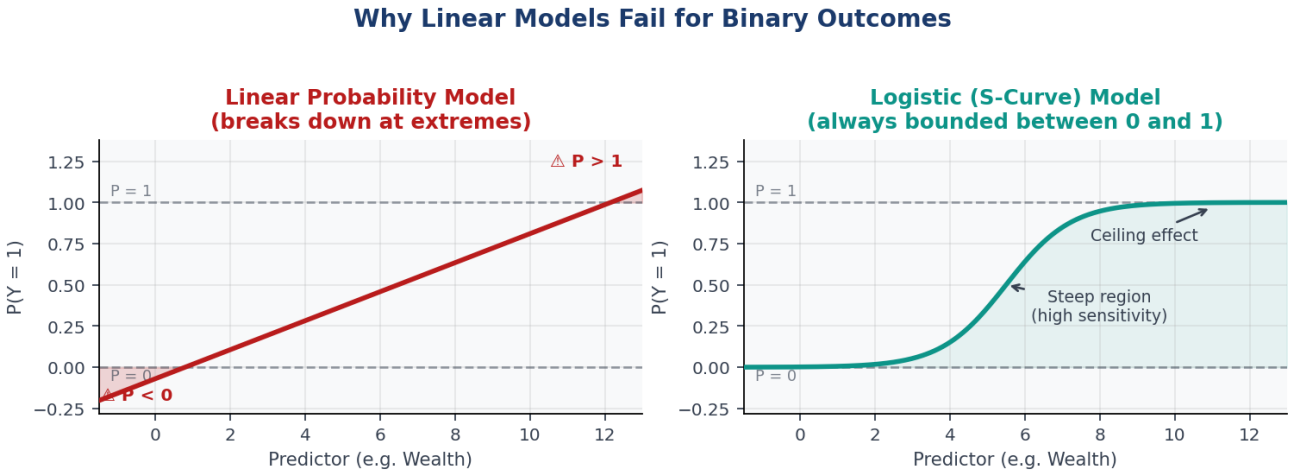

Figure 2 — The Linear Probability Model can yield predicted probabilities outside [0, 1] (statistically embarrassing). The logistic model always stays bounded by the S-curve.

A Real-World Horror Story (Fairlie, 2017)

Imagine looking at the gap in computer ownership between White Californians and Latino Oklahomans. The gap in average house values between these two groups is massive ($355k vs. $49k). If you use a linear BO decomposition, the formula says:

Contribution of House Value = $(\bar{X}_{\text{House,W}} - \bar{X}_{\text{House,L}}) \times \beta_{\text{House}}$

Because the X-difference is so huge, the linear model might tell you that house values explain **140 % of the gap.** That is statistically embarrassing.

Expressed as a formula, the problematic term is:

$$\text{Contribution} = (\bar{X}^{W}_{HV} - \bar{X}^{L}_{HV})\,\hat{\beta}_{HV}$$

The Fairlie Solution

Fairlie (1999, 2005) solved this by pushing the decomposition through the non-linear function (let's use the logistic CDF, denoted F). Instead of multiplying averages, we calculate the predicted probability for *every single person* in the sample, and then take the average. The total gap $\Delta\bar{Y}$ becomes:

$$\Delta\bar{Y} = \underbrace{\left[\frac{1}{N^W}\sum_{i=1}^{N^W} F(X_i^W\hat{\beta}^W) - \frac{1}{N^B}\sum_{j=1}^{N^B} F(X_j^B\hat{\beta}^W)\right]}_{\text{Explained}} + \underbrace{\left[\frac{1}{N^B}\sum_{j=1}^{N^B} F(X_j^B\hat{\beta}^W) - \frac{1}{N^B}\sum_{j=1}^{N^B} F(X_j^B\hat{\beta}^B)\right]}_{\text{Unexplained}}$$

- **First bracket — Explained:** Compares White characteristics (X^W) to Black characteristics (X^B), holding the coefficient vector constant at β^W.
- **Second bracket — Unexplained:** Holds Black characteristics fixed but changes the rulebook from White coefficients to Black coefficients. (Pooled coefficients β^* are typically used instead of β^W, just as in the linear case.)

The Swapping Algorithm

Getting the *total* explained gap is easy. But how do we isolate the contribution of just *one* variable, like Education (X_1)?

In OLS, we just multiplied the difference in means by the coefficient. In Logit, we can't do that because the variables are mashed together inside the $F(\cdot)$ function. The contribution of Education depends on what your Age and Income are.

Fairlie's genius solution is a **swapping algorithm**. Assume we have an equal number of people in Group A and Group B, matched one-to-one. If we want to find the contribution of Education (X_1), we take Group B, but we *swap out* their education levels for their matched partners in Group A, while keeping all their other traits (X_2) exactly the same. Formally:

$$\text{Contrib.}(X_1) = \frac{1}{N^B}\sum_{i}^{N^B}\left[F\left(\hat{\alpha}^* + \hat{\beta}_1^* X_{1i}^A + \hat{\beta}_2^* X_{2i}^B\right) - F\left(\hat{\alpha}^* + \hat{\beta}_1^* X_{1i}^B + \hat{\beta}_2^* X_{2i}^B\right)\right]$$

You are literally simulating a universe where Group B suddenly magically possesses Group A's education, and you calculate how much their average predicted probability goes up. That increase is the contribution of Education!

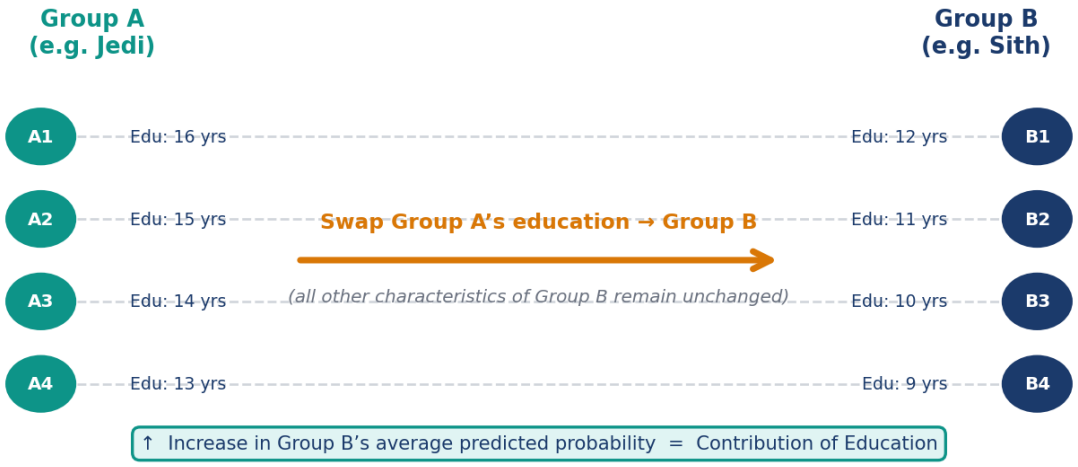

Figure 3 — The swapping algorithm in action. Each Group B individual gets Group A's education level while keeping all other characteristics unchanged. The resulting increase in average predicted probability is the contribution of education to the gap.

Handling Different Sample Sizes

What if Group A (Whites) has 10,000 people and Group B (Minorities) has 500 people? You can't match them one-to-one. Fairlie's algorithm handles this via **random subsampling**:

1. Draw a random sample of 500 Whites from the 10,000.
2. Randomly pair them up with the 500 Minorities.
3. Do the swapping math.
4. Because random draws can be weird, repeat this process 1,000 times and take the average of the results.

> **Historical Note**
>
> In Fairlie's original 1999 paper, he ranked people by their predicted probabilities and matched them highest-to-highest. In 2007, he revised this to match them **completely randomly**, which better preserves the true distribution of the data.

4. The Ghost of Path Dependence (Order Matters!)

Here is the one terrifying caveat about the Fairlie non-linear decomposition: **Path Dependence**.

Because the logistic function is an S-curve, moving your position on the curve changes how much a variable impacts the outcome. Imagine you are trying to calculate the contributions of Income (X_1) and Education (X_2).

- If you swap Income *first*, it might push the individual from a 10 % probability to a 50 % probability (the steepest part of the S-curve). Then, swapping Education pushes them from 50 % to 60 %.
- If you swap Education *first*, Education gets the steep part of the curve, and Income gets the flatter top part.

The order in which you swap the variables changes their individual contribution estimates! If you just pick an arbitrary order in your code, your results are technically biased by your aesthetic choices.

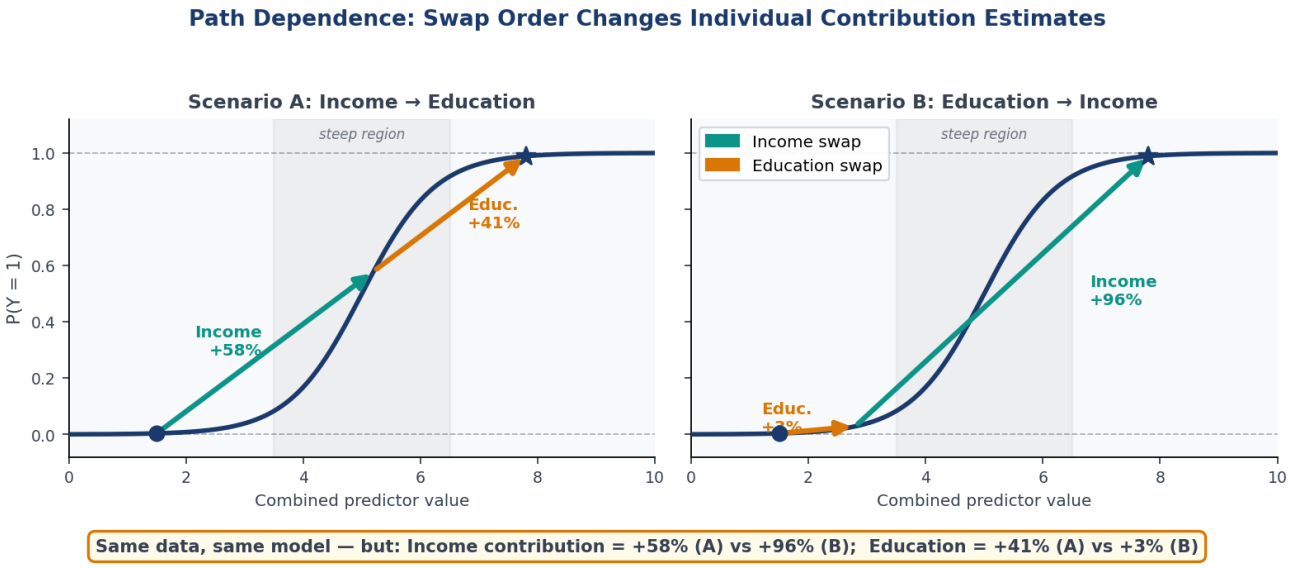
Path Dependence: Swap Order Changes Individual Contribution Estimates
Scenario A: Income → Education
Scenario B: Education → Income
steep region
Income swap
Education swap
Income +58%
Educ. +41%
Income +96%
Educ. +3%
P(Y = 1)
Combined predictor value
Same data, same model — but: Income contribution = +58% (A) vs +96% (B); Education = +41% (A) vs +3% (B)

Figure 4 — Two scenarios with identical data. In Scenario A, Income is swapped first and captures the steep region of the S-curve (large contribution); Education is left with the flat upper portion (small contribution). Swap the order in Scenario B and the attribution reverses. The total gap is the same — only the breakdown differs.

The Solution: Randomize the Order (Fairlie, 2017)

Since we are already running 1,000 replications to handle the different sample sizes, we simply tell the computer to **randomly shuffle the order in which the variables are swapped** for every single replication. Across 1,000 runs, every possible permutation of variable ordering will be tested. By taking the average across all replications, the path dependence washes out entirely. You get the true, unbiased average contribution of each variable. Crisis averted.

5. Survey Weights: The Final Boss

If you are using data from the Census, the Current Population Survey (CPS), or any major panel, you are going to have complex survey weights. Some individuals represent 100 people in the population, and others represent 5,000. If you just run the standard matching algorithm, you are treating everyone as equal, which violates the survey design.

The Weighted Swapping Method (Jann 2006; Fairlie 2017) handles this as follows:

1. You draw the Group A subsample **with replacement**. The probability of an individual being drawn is **proportional to their survey weight**. (A person with a weight of 5,000 is 50 times more likely to be drawn into the matching sample than someone with a weight of 100.)
2. Do the exact same proportional drawing for Group B! Create a new sample of Group B drawn with replacement based on their weights, equal in size to the Group A draw.
3. Perform the matching and swapping on these weight-proportional samples.

By ensuring both sides of the matching process are heavily populated by the highest-weighted individuals, your final decomposition correctly scales to the national population.

Summary

1	**Blinder-Oaxaca** breaks mean gaps down into **Endowments (Explained)** and **Coefficients (Unexplained/Bias)**.
2	If you want the full systemic picture, compare **Structural models** (current controls) with **Reduced Form models** (exogenous “circumstances of birth” controls).
3	**The Index Number Problem** asks whose baseline coefficients to use. Usually, pooling the sample (Neumark) is the safest bet — but don’t forget the group dummy variable!

4	If you have a **binary outcome**, linear probability models will ruin your life if group differences in X are large. Use the **Fairlie Non-Linear Decomposition**.
5	Fairlie calculates contributions by **swapping** characteristics between groups.
6	Because non-linear models are S-curves, swapping is subject to **Path Dependence**. Fix this by running 1,000+ replications with **randomized variable ordering**.
7	To use **Survey Weights** in Fairlie, draw your matching subsamples *with replacement* proportional to the weights.

Decomposition methods are incredibly powerful tools. They take the blunt instrument of "Group A wins, Group B loses" and turn it into a scalpel, allowing you to dissect exactly where the disparities are coming from. Now get out there, boot up R or Stata, and mind the gap.

References

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources, 8*(4), 436–455.

Cotton, J. (1988). On the decomposition of wage differentials. *The Review of Economics and Statistics, 70*(2), 236–243.

Fairlie, R. W. (1999). The absence of the African-American owned business: An analysis of the dynamics of self-employment. *Journal of Labor Economics, 17*(1), 80–108.

Fairlie, R. W. (2005). An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement, 30*(4), 305–316.

Fairlie, R. W. (2017). *Addressing path dependence and incorporating sample weights in the nonlinear Blinder-Oaxaca decomposition technique for logit, probit and other nonlinear models* (SIEPR Discussion Paper No. 17-013). Stanford Institute for Economic Policy Research.

Gardeazabal, J., & Ugidos, A. (2004). More on identification in detailed wage decompositions. *The Review of Economics and Statistics, 86*(4), 1034–1036.

Jann, B. (2006). *fairlie*: Stata module to generate nonlinear decomposition of binary outcome differentials. *Statistical Software Components.* http://ideas.repec.org/c/boc/bocode/s456727.html

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *The Journal of Human Resources, 23*(3), 279–295.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review, 14*(3), 693–709.

Oaxaca, R. L., & Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics, 61*(1), 5–21.

Reimers, C. W. (1983). Labor market discrimination against Hispanic and black men. *The Review of Economics and Statistics, 65*(4), 570–579.

Yun, M.-S. (2005). A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry, 43*(4), 766–772.